

# Robust Group Fused Lasso for Multisample CNV Detection under Uncertainty

Hossein Sharifi Noghabi<sup>\*a,b</sup>, Majid Mohammadi<sup>c</sup>

<sup>a</sup>*Department of Computer Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran*

<sup>b</sup>*Center of Excellence on Soft Computing and Intelligent Information Processing (SCIIP)*

<sup>c</sup>*Faculty of Technology, Policy and Management, Delft University of Technology, Netherlands*

---

## Abstract

One of the most important needs in the post-genome era is providing the researchers with reliable and efficient computational tools to extract and analyze this huge amount of biological data, in which DNA copy number variation (CNV) is a vitally important one. Array-based comparative genomic hybridization (aCGH) is a common approach in order to detect CNVs. Most of methods for this purpose were proposed for one-dimensional profile. However, slightly this focus has moved from one- to multi-dimensional signals. In addition, since contamination of these profiles with noise is always an issue, it is highly important to have a robust method for analyzing multi-sample aCGH data. In this paper, we propose Robust Grouped Fused Lasso (RGFL) which utilizes the Robust Group Total Variations (RGTV). Instead of  $l_{2,1}$  norm, the  $l_1$ - $l_2$  M-estimator is used which is more robust in dealing with non-Gaussian noise and high corruption. More importantly, Correntropy (Welsch M-estimator) is also applied for fitting error. Extensive experiments indicate that the proposed method outperforms the state-of-the art algorithms and techniques under a wide range of scenarios with diverse noises.

*Keywords:* Cancer, CNV, aCGH, half quadratic, M-estimators

---

## 1. Introduction

Determination of changes in one- or multi-dimensional signals is of great importance and significance in fields such as Artificial Intelligence [1], Computer Networks [2] and Economy [3]. In addition to aforementioned areas, one- or multi-dimensional signals analysis has also found its way in biology and medicine [4]. Finding and specifying the place/time when a change is happening is significantly important especially when it comes to Copy-Number Variations (CNVs) in the genome [5]. The technique of microarray comparative genomic hybridization (arrayCGH or aCGH) paves the way for one to investigate genome-wide in order to capture all of these alternations [25]. Most of aCGH analysis methods were proposed for one-dimensional profile. However, slightly this focus has moved from one- to multi-dimensional signals where research are capable of capturing alternations from multi profiles simultaneously [7, 8]. Although multi-dimensional analysis brought more insight and information about aCGH landscape, most recently group approaches have been proposed for analyzing aCGH profiles [9]. There are two main advantages for group approach: 1) since these methods are considering a group of multi-dimensional profiles, they can provide us with more information about the structure of the profiles (where CNVs are occurring) and 2) this approach tends to be faster than previous methods as all profiles are considered at the same time [10]. Among all of the proposed methods for change point detection, Lasso [11] has found its way in numerous area from acoustics and speech processing [12] to compress sensing [13] and biomedical data analysis especially aCGH profiles [9]. Tibshirani et al. [14] proposed spatial smoothing and hot spot detection for CGH data using the fused lasso for one-dimensional aCGH analysis. Nowak et al. [5] proposed fused lasso latent feature model for multi-dimensional profiles.

Both of these methods applied the TV (Total Variation) norm for each sample independently and assume that the aCGH matrix is sparse. However, Zhou et al. [26, 25] proposed piecewise-constant and low-rank

---

\*Corresponding Author

approximation (PLA) which assume this matrix is low-rank. Similar to previous methods, PLA is also take advantage of TV over for each sample. However, in order to deal with multi-dimensional features from a group perspective, Alvaiz et al. [10] proposed Group Fused Lasso (GFL) that utilizes Group Total Variation (GTV) for regularization. GTV uses the  $l_{2,1}$  norm instead of the  $l_1$ . Bleakley et al. [9] utilized GFL with a weighted approach for multiple change point detection in aCGH profiles which lead to piecewise-constant approximations as well. While most methods are typically using first-order or alternating minimization schemes, Wytock et al. [16] proposed a fast Newton methods for the GFL that uses combination of a projected Newton method and a primal active set approach which is substantially fast.

### 1.1. Contribution

Based on  $l_1$ - $l_2$  and Welsch M-estimators, Robust Group Fused Lasso (RGFL) is proposed to recover true aCGH profile from noisy measurement. The proposed problem for RGFL is not convex, but its efficient solution is given by half quadratic (HQ) programming and its convergence is guaranteed. Due to the smoothness of the proposed algorithm, it is significantly faster than the other state-of-the-art algorithms which usually solve non-smooth problems. Extensive experimental results on simulated and real data illustrate the robustness of RGFL, especially when data are highly corrupted with non-Gaussian noise.

## 2. Robust Group Fused Lasso

### 2.1. Correntropy

To deal with non-Gaussian noise and impulsive noise, the concept of Correntropy is proposed in the realm of signal processing [17, 18]. Correntropy of two arbitrary variable  $X$  and  $Y$  is defined as

$$V(X, Y) = E[\mathcal{K}(X - Y)] \quad (1)$$

where  $E[\cdot]$  is expectation operator and  $\mathcal{K}$  is a kernel function.

In practice, only finite number of samples  $\{(X_i, Y_i)\}_{i=1}^n$  are available and the joint probability distribution of  $X$  and  $Y$  is usually unknown. Thus, Eq. (1) can be estimated as

$$V(X, Y) = \frac{1}{N} \sum_{i=1}^n \mathcal{K}_\sigma(X_i - Y_i) \quad (2)$$

where  $\mathcal{K}_\sigma$  is Gaussian kernel function with the width of  $\sigma$  and  $\mathcal{K}_\sigma(x, y) = \exp(-\frac{x^2}{2\sigma^2})$ . Correntropy is used to measure the similarity between  $X$  and  $Y$ . Liu et al. [17] induced a metric based on (2) for any two discrete vectors. They introduced the Correntropy induced metric (CIM) for any two vectors  $x \in R^n$  and  $y \in R^n$  as [18]

$$CIM(x, y) = \left( \mathcal{K}(0) - \frac{1}{n} \sum_{i=1}^n \mathcal{K}_\sigma(x_i - y_i) \right)^{\frac{1}{2}}. \quad (3)$$

In contrary to mean square error (MSE), it should be noticed that Correntropy is a local similarity measure and the value of Correntropy is highly related to the width of Gaussian kernel function [17].

Correntropy can be applied in any noise environment because it has the probabilistic meaning of maximizing the error density function at the origin [17]. Further, if we choose Gaussian kernel function, Correntropy is a Welsch M-estimator in robust statistics [19].

### 2.2. Proposed formulation

Given a dataset  $D \in R^{n \times m}$  of aCGH profiles, it is desired to find the decomposition

$$D = B + E \quad (4)$$

where  $B \in R^{n \times m}$  and  $E \in R^{n \times m}$  are true aCGH profiles and an additive arbitrary noise, respectively. A main problem in finding decomposition similar to (4) is that aCGH data are highly corrupted with various types of noise. Hence, the robust group fused lasso (RGFL) is proposed as follow

$$\min_B \sum_{i,j} \phi_c(D_{ij} - B_{ij}) + \alpha \sum_{i,j} \phi_r(B_{ij}) + \gamma \sum_{i,j} \phi_r((TB)_{ij}) \quad (5)$$

where  $\alpha, \gamma > 0$  are regularization parameters,  $\phi_c(x) = 1 - \exp(-\frac{x^2}{\sigma^2})$  and  $\phi_r(x) = \sqrt{x^2 + \epsilon}$  which is  $l_1 - l_2$  loss function in M-estimation and  $T \in R^{n-1 \times n}$  is

$$T = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{bmatrix}$$

which makes  $TB$  takes the differences of the columns of  $B$ .

In fact,  $\phi_c(\cdot)$  in Eq. (5) is a robust substitution of  $l_1 - l_2$  in group fused lasso [20]. Although  $l_1$  is convex, but it is not differentiable at zero point and it results in oscillating around optimum and converging slowly toward the optimal solution. In contrast,  $l_1 - l_2$  is differentiable around zero point and can be solved significantly faster. Further, when  $\alpha \rightarrow 0$  the  $l_1 - l_2$  loss is equivalent to  $l_1$  loss. Hence, The second term in Eq. (5) forces the desired matrix  $B$  to be group sparse and the third term is a robust group total variation (RGTV). The first term is also CIM which is a robust measure for fitting error and can deal with diverse noise.

Table 1: HQ loss functions and their corresponding minimizer functions

### 3. Solution of the proposed formulation

An iterative procedure to find the optimal solution of the proposed problem (5) is given by HQ minimization and the convergence of the solution is analyzed. Then, the optimal parameters for the proposed problem is given.

#### 3.1. Half Quadratic Minimization

The proposed formulation (5) is not convex and cannot be solved directly. Fortunately, HQ minimization can be utilized to find the efficient optimal solution of Eq. (5) by an alternating procedure. Based on the conjugate function theory [21] and HQ theory [22, 23], the following lemma can be obtained.

**Lemma 3.1.** *Let  $\phi(x)$  is a loss function which satisfies the five conditions of additive HQ listed in [24]. Then, for any fixed  $x$*

$$\phi(x) = \inf_{s \in R} \{(x - s)^2 + \psi(s)\} \quad (6)$$

where  $s$  is an auxiliary variable which is determined by the minimizer function  $\delta(\cdot)$  related only to  $\phi(\cdot)$  (see Table 1).

Replacing  $\phi(\cdot)$  in Eq. (5) by Eq. (6), we obtain the following augmented cost function

$$\mathcal{J}(B, W, E, M) = \sum_{i,j} (D_{ij} - B_{ij} - E_{ij})^2 + \psi_c(E_{ij}) + \alpha \sum_{i,j} (B_{ij} - W_{ij})^2 + \psi_r(W_{ij}) + \gamma \sum_{i,j} ((TB)_{ij} - M_{ij})^2 + \psi_r(M_{ij}) \quad (7)$$

where  $E, W$  and  $M$  are HQ auxiliary variables. According to HQ,  $\mathcal{J}$  can be optimized by alternating minimization procedure. The HQ auxiliary variables are only related to the minimizer functions. Thus  $\sum \psi(E_{ij})$ ,  $\sum \psi(W_{ij})$  and  $\sum \psi(M_{ij})$  become fixed and can be ignored for updating  $B$ :

$$B = \arg \min_B \|D - B - E\|_F^2 + \alpha \|B - W\|_F^2 + \gamma \|TB - M\|_F^2.$$

By derivating the above problem by  $B$ , we obtain

$$\begin{aligned} B + E - D + \alpha(B - W) + \gamma T^T(TB - M) &= 0 \\ B &= (\gamma T^T T + (\alpha + 1)I)^{-1}(D - E + \alpha W + \gamma T^T M) \end{aligned} \quad (8)$$

where  $I$  is the identity matrix. In short, the following steps must be taken for optimizing augmented cost function (7):

$$\begin{aligned}
 E_{ij}^{k+1} &= \delta_c(D_{ij} - B_{ij}^k) \\
 W_{ij}^{k+1} &= \delta_r(B_{ij}^k) \\
 M_{ij}^{k+1} &= \delta_r((TB^k)_{ij}) \\
 B^{k+1} &= (\gamma T^T T + (\alpha + 1)I)^{-1}(D - E^{k+1} + \alpha W^{k+1} + \gamma T^T M^{k+1})
 \end{aligned} \tag{9}$$

Algorithm 1 summarizes the procedure of RGFL.

---

**Algorithm 1** Robust Group Fused Lasso

---

**Require:**  $D$

- 1: *Initialize:*  $D \in R^{m \times n}$ ,  $\hat{\mu} = \theta\mu_0$ ,  $B_0 = B_{-1} = 0$ ,  $E_0 = E_{-1} = 0$ ,  $Z_0 = Z_{-1} = 0$ ,  $Y = 0$ ,  $t_0 = t_{-1} = 1$ ,
  - 2: **while** not converge **do**
  - 3:      $E_{ij}^{k+1} = \delta_c(D_{ij} - B_{ij}^k)$
  - 4:      $W_{ij}^{k+1} = \delta_r(B_{ij}^k)$
  - 5:      $M_{ij}^{k+1} = \delta_r((TB^k)_{ij})$
  - 6:      $B^{k+1} = (\gamma T^T T + (\alpha + 1)I)^{-1}(D - E^{k+1} + \alpha W^{k+1} + \gamma T^T M^{k+1})$
  - end**
  - 7: **return**  $B$
- 

### 3.2. Convergence Analysis

For fixed  $E$ ,  $M$  and  $W$ , it is readily seen that

$$J(B^{k+1}, E^k, W^k, M^k) \leq J(B^k, E^k, W^k, M^k)$$

Further, according to the properties of HQ minimizer function  $\delta(\cdot)$  [24, 23] for a fixed  $B$

$$J(B^{k+1}, E^{k+1}, W^{k+1}, M^{k+1}) \leq J(B^{k+1}, E^k, W^k, M^k). \tag{10}$$

As Correntropy is bounded [17, 18], the sequence

$$\{\dots, J(B^{k+1}, E^{k+1}, W^{k+1}, M^{k+1}), J(B^{k+1}, E^k, W^k, M^k), J(B^k, E^k, W^k, M^k), \dots\}$$

converges as  $k \rightarrow \infty$ .

The inverse problem (8) can be solved using iterative method for linear equation system  $Ax = b$  that can even accelerate the convergence.

### 3.3. Parameter Selection

There are two parameters  $\alpha$  and  $\gamma$  which can dramatically influence the performance of the proposed algorithm. In this section we go through a procedure by which these parameters are obtained. To do so, the data are divided into two subsets:  $S_1$  and  $S_2$ . Let  $P_\Omega(\cdot)$  be the projection operator defined as

$$P_\Omega(M)(i, j) = \begin{cases} M_{ij} & \text{if } (i, j) \in \Omega \\ 0 & \text{otherwise} \end{cases}. \tag{11}$$

First, we perform the proposed method on the subset  $S_1$  with different values for  $\alpha$  and  $\gamma$ , i.e.

$$\min_B \sum_{i,j} \phi_c(P_{S_1}(D_{ij}) - B_{ij}) + \alpha \sum_{i,j} \phi_r(B_{ij}) + \gamma \sum_{i,j} \phi_r((TB)_{ij}). \tag{12}$$

The solution for above problem is called  $\hat{B}(\alpha, \gamma)$  for each  $(\alpha, \gamma)$  pair. Then, the pair is selected which minimize the prediction error defined as [25]

$$Err(\alpha, \gamma) = \frac{1}{2} \|P_{S_2}(D) - \hat{B}(\alpha, \gamma)\|_F^2.$$

Table 2: Probability density functions (PDF) and their parameters utilized as an additive noise to simulated data.

Figure 1: ROC for simulated data with non-Gaussian noise and different SNRs and shared percentages. Each row and each column are dedicated to a specific SNR and share percentage, respectively. The  $x$ -axis and  $y$  are false discovery rate (FDR) and true discovery rate (TDR), respectively.

## 4. Experimental Results

In order to evaluate the proposed method, we performed our experiments on simulated and real data sets.

### 4.1. Simulated Data

In this section, the performance of the proposed method is compared with four well known methods including, TVSP [25], PLA [26], RCLR [8] and GFLseg [20]. In each case, 50 samples with 500 probes are generated according to the model (4) and various types of noise are added for investigation of the robustness. These noises are Uniform, Gamma, Beta and Laplace with settings tabulated in Table 2. Similar to [25], two types of aberrations are considered: the first type is added to each sample individually and the second type is shared among all samples at the same locations. The ratio between shared and total aberrations is denoted as *shared percentage*.

For each of these methods the Matlab implementations of them are downloaded and the parameters are tuned accordingly. Figure 1 illustrates the receiver operating characteristic (ROC) curves under Gamma, Uniform and Laplace noises and Figure 2 presents the comparison between stated methods in the case of accuracy under Beta noise.

Obviously, more deviation from the diagonal line in ROC indicates better performance for the corresponding method. As plotted in this figure, the proposed method consistently and significantly outperforms PLA, GFLSeq and TVSP in all of the experimented noises. Especially for the higher SNRs and shared percentages which indicates that applying M-estimators and Correntropy makes the proposed method significantly robust against all of the studied noises and as a result much better performance in detection of common aberrations among multi-samples. As shown in Figure 1, RCLR seems to be extremely competitive, the reason for this observation is that similar to the proposed method, RCLR is also take advantage of Correntropy which makes it robust against uncertainty. However, despite this advantage of RCLR, the proposed method significantly and consistently outperforms it in Gamma and Uniform noises under both high and low SNR and shared percentage. This means utilizing Correntropy and M-estimators simultaneously, provides the proposed method with more robustness against uncertainty. In the case of Laplace noise with high value for SNR, RCLR appears to be slightly better than the proposed method, however, for lower SNR the proposed method again outperforms the RCLR.

These conclusions and observations are also deducible in Figure 2 for the accuracy under Beta noise. As illustrated in this figure, the proposed method significantly outperforms PLA, GFLSeq and TVSP in both high and low SNRs and shared percentages and gets extremely competitive with RCLR in higher SNR values. However, the proposed method achieves better accuracy for both high and low shared percentages when SNR is lower in comparison with RCLR.

Figure 2: The comparison on simulated data from accuracy point of view. The data in this experiment are contaminated with Beta noise (see Table 2).

### 4.2. Real Data

We investigate the performance of the proposed method on two independent real data sets about breast cancer. The Chin data set [27] includes profiles of 2,149 DNA clones for 141 primary breast tumors and the Pollack data set [28] has 6,691 human mapped genes for 44 advanced primary breast tumors. The results of these data sets are brought in Figure 3. The recovered profiles are plotted by heat map on the top, the recovered profile for a random sample in the middle and bar diagram, which presents the sum of number of gains over all samples for each probe with a threshold equals to one, at the bottom.

In this figure, discovered regions with high amplifications are highly probable to be important for breast cancer, i.e., probes 67-70 and 39-42 are discovered for the Chin data set [27] and probes 178-184 are for the Pollack data set [28]. Interestingly, several identified functionally crucial regions for breast cancer, i.e., transcription regulation protein PPARBP, the receptor tyrosine kinase ERBB2 and the adaptor protein GRB7 are located within the discovered regions via the proposed method. Moreover, as illustrated in this figure  $\log_2$ -ratio of the denoised signals for both Chin and Pollack data sets are significantly smooth.

Figure 3: Heat and bar diagram for real datasets. (a) Performing RGFL on dataset introduced in [28] (b) Performing RGFL on dataset introduced in [27].

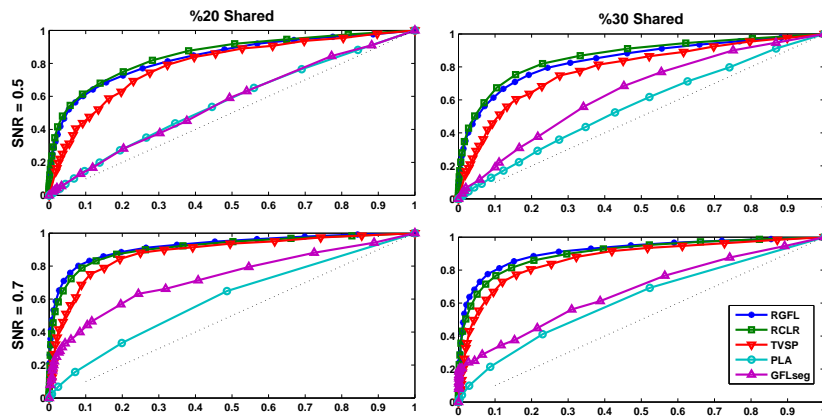
## 5. Conclusion

This paper presented Robust Group Fused Lasso (RGFL) to detect multiple changepoints in multi-dimensional signals. The proposed problem in this paper was non-convex, but an iterative procedure to find its efficient solution was given by HQ programming and its convergence analysis was given. In comparison to other state-of-the-art algorithms, the proposed method has shown significant robustness in dealing with high corruption, especially when data are mixed with non-Gaussian noise.

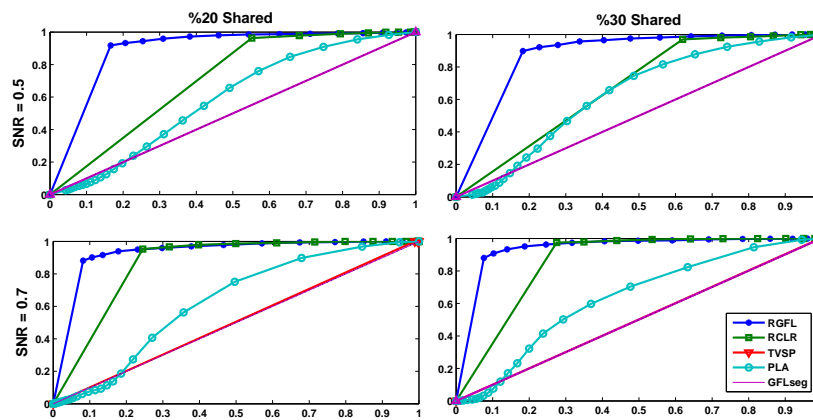
## 6. Reference

- [1] F. Desobry, M. Davy, and C. Doncarli, "An online kernel change detection algorithm," *Signal Processing, IEEE Transactions on*, vol. 53, pp. 2961–2974, Aug 2005.
- [2] A. Tartakovsky, B. Rozovskii, R. Blazek, and H. Kim, "A novel approach to detection of intrusions in computer networks via adaptive sequential and batch-sequential change-point detection methods," *Signal Processing, IEEE Transactions on*, vol. 54, pp. 3372–3382, Sept 2006.
- [3] M. Talih and N. Hengartner, "Structural learning with time-varying components: tracking the cross-section of the financial time series," *J. Royal Statist. Soc. B*, pp. 321–341, 2005.
- [4] N. R. Zhang, D. O. Siegmund, H. Ji, and J. Z. Li, "Detecting simultaneous changepoints in multiple sequences," *Biometrika*, vol. 97, no. 3, pp. 631–645, 2010.
- [5] G. Nowak, T. Hastie, J. R. Pollack, and R. Tibshirani, "A fused lasso latent feature model for analyzing multi-sample acgh data," *Biostatistics*, p. kxr012, 2011.
- [6] X. Zhou, C. Yang, X. Wan, H. Zhao, and W. Yu, "Multisample acgh data analysis via total variation and spectral regularization," *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, vol. 10, pp. 230–235, Jan. 2013.
- [7] Z. Tian, H. Zhang, and R. Kuang, "Sparse group selection on fused lasso components for identifying group-specific dna copy number variations," in *Proceedings of the 2012 IEEE 12th International Conference on Data Mining, ICDM '12*, pp. 665–674, 2012.
- [8] M. Mohammadi, G. A. Hodtani, and M. Yassi, "A robust correntropy-based method for analyzing multisample acgh data," *Genomics*, 2015.
- [9] K. Bleakley and J.-P. Vert, "The group fused lasso for multiple change-point detection," *arXiv preprint arXiv:1106.4199*, 2011.
- [10] C. M. Alaíz, Á. Barbero, and J. R. Dorransoro, "Group fused lasso," in *Artificial Neural Networks and Machine Learning—ICANN 2013*, pp. 66–73, Springer, 2013.
- [11] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.
- [12] A. Gibberd and J. Nelson, "High dimensional changepoint detection with a dynamic graphical lasso," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 2684–2688, May 2014.
- [13] D. Angelosante, G. Giannakis, and E. Grossi, "Compressed sensing of time-varying signals," in *Digital Signal Processing, 2009 16th International Conference on*, pp. 1–8, July 2009.
- [14] R. Tibshirani and P. Wang, "Spatial smoothing and hot spot detection for cgh data using the fused lasso," *Biostatistics*, vol. 9, no. 1, pp. 18–29, 2008.
- [15] X. Zhou, J. Liu, X. Wan, and W. Yu, "Piecewise-constant and low-rank approximation for identification of recurrent copy number variations," *Bioinformatics*, vol. 30, no. 14, pp. 1943–1949, 2014.
- [16] M. Wytoczek, S. Sra, and J. Z. Kolter, "Fast newton methods for the group fused lasso," in *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, 2014.
- [17] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: properties and applications in non-gaussian signal processing," *Signal Processing, IEEE Transactions on*, vol. 55, no. 11, pp. 5286–5298, 2007.
- [18] I. Santamaría, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: definition, properties, and application to blind equalization," *Signal Processing, IEEE Transactions on*, vol. 54, no. 6, pp. 2187–2197, 2006.
- [19] P. J. Huber, *Robust statistics*. Springer, 2011.
- [20] K. Bleakley and J.-P. Vert, "The group fused lasso for multiple change-point detection," *arXiv preprint arXiv:1106.4199*, 2011.
- [21] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [22] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 3, pp. 367–383, 1992.

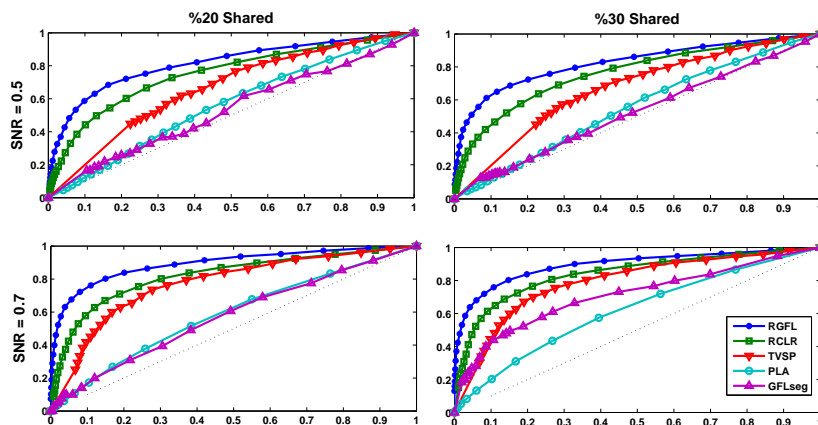
- [23] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *Image Processing, IEEE Transactions on*, vol. 4, no. 7, pp. 932–946, 1995.
- [24] M. Nikolova and M. K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM Journal on Scientific computing*, vol. 27, no. 3, pp. 937–966, 2005.
- [25] X. Zhou, C. Yang, X. Wan, H. Zhao, and W. Yu, "Multisample acgh data analysis via total variation and spectral regularization," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 10, no. 1, pp. 230–235, 2013.
- [26] X. Zhou, J. Liu, X. Wan, and W. Yu, "Piecewise-constant and low-rank approximation for identification of recurrent copy number variations," *Bioinformatics*, p. btu131, 2014.
- [27] K. Chin, S. DeVries, J. Fridlyand, P. T. Spellman, R. Roydasgupta, W.-L. Kuo, A. Lapuk, R. M. Neve, Z. Qian, T. Ryder, *et al.*, "Genomic and transcriptional aberrations linked to breast cancer pathophysiologies," *Cancer cell*, vol. 10, no. 6, pp. 529–541, 2006.
- [28] J. R. Pollack, T. Sørlie, C. M. Perou, C. A. Rees, S. S. Jeffrey, P. E. Lonning, R. Tibshirani, D. Botstein, A.-L. Børresen-Dale, and P. O. Brown, "Microarray analysis reveals a major direct role of dna copy number alteration in the transcriptional program of human breast tumors," *Proceedings of the National Academy of Sciences*, vol. 99, no. 20, pp. 12963–12968, 2002.



(a) Data contaminated with Laplace noise



(b) Data contaminated with Gamma noise



(c) Data contaminated with Uniform noise

Figure 1: ROC for simulated data with non-Gaussian noise and different SNRs and shared percentages. Each row and each column are dedicated to a specific SNR and share percentage, respectively. The  $x$ -axis and  $y$  are false discovery rate (FDR) and true discovery rate (TDR), respectively.



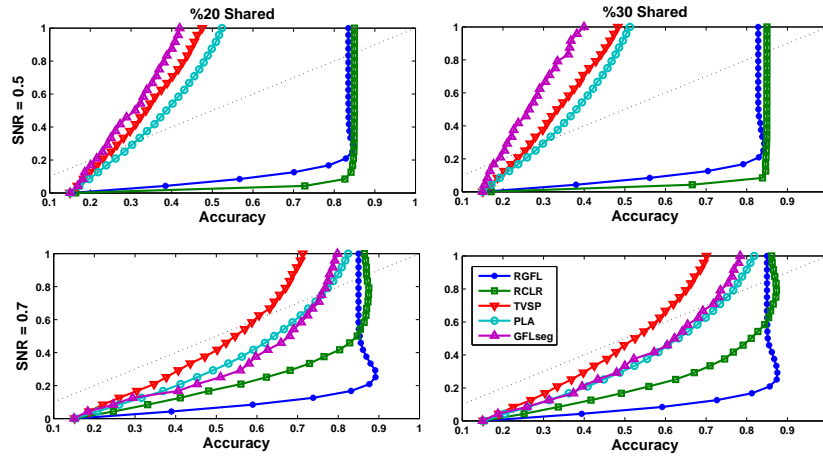
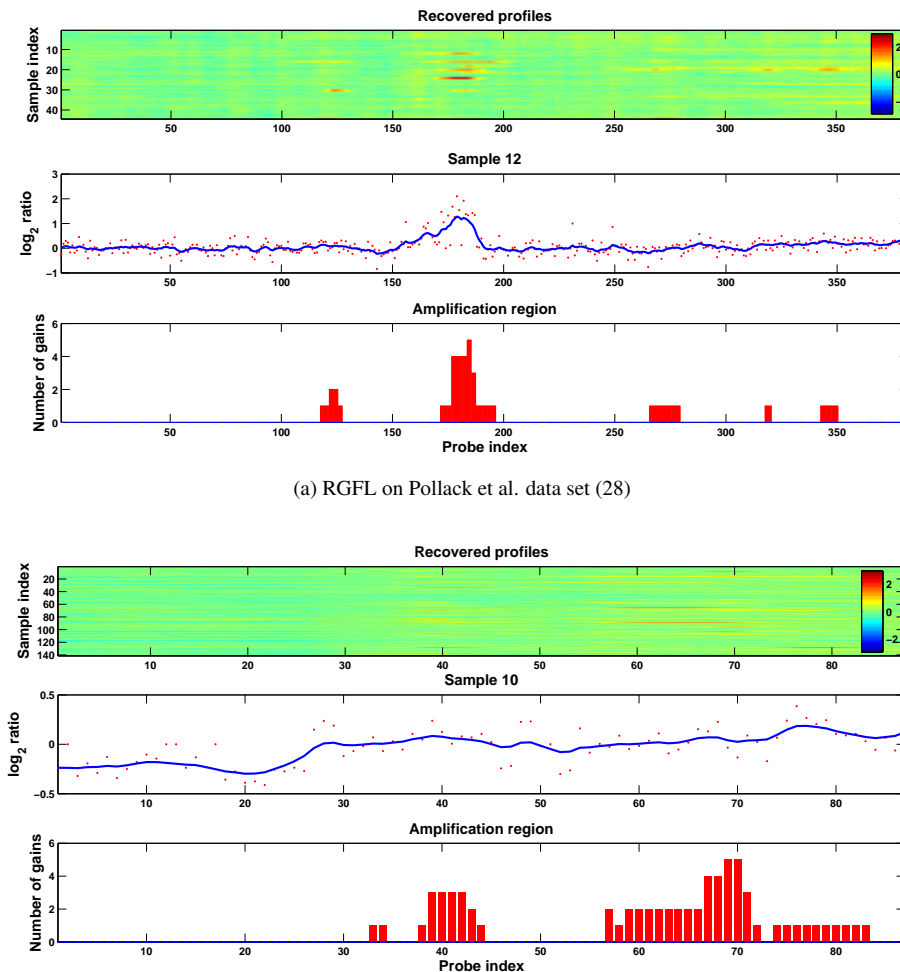


Figure 2: The comparison on simulated data from accuracy point of view. The data in this experiment are contaminated with Beta noise (see Table 1).



(a) RGFL on Pollack et al. data set (28)

(b) RGFL on Chin et al. data set (27)

Figure 3: Heat and bar diagram for real datasets. (a) Performing RGFL on dataset introduced in (28) (b) Performing RGFL on dataset introduced in (27).

Table 1: HQ loss functions and their corresponding minimizer functions

estimators	11-12	Welsch
loss function	$\sqrt{\alpha + t^2} - 1$	$1 - \exp(-\frac{t^2}{\sigma^2})$
Additive form $\delta(t)$	$t - \frac{t}{\sqrt{\alpha+t^2}}$	$t - t \exp(-\frac{t^2}{\sigma^2})$

Table 2: Probability density functions (PDF) and their paramters utilized as an additive noise to simulated data.

distribution	Laplace	Gamma	Uniform	Beta
PDF	$f_{Lap}(\mu, \sigma) = \frac{1}{2\sigma} e^{-\frac{ x-\mu }{\sigma}}$	$f_{\gamma}(k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\theta^k \Gamma(k)}$	$\begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & otherwise \end{cases}$	$f_{\beta}(\alpha, \beta) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)}$
settings	$\mu = 0, \sigma = 1$	$k = \theta = 1$	$a = 1 - \sqrt{3}, b = 1 + \sqrt{3}$	$\alpha = \beta = 1$